

I hereby certify that this correspondence is being transmitted by facsimile (703-872-9306) to:

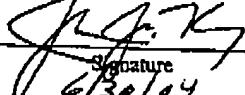
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

on June 30, 2004

Date of Deposit

John J. King

Name of applicant, assignee or  
Registered Representative



Signature

6/30/04

Date of Signature

Our Case No. 10566/4

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:	)
Eliot R. Long	)
Serial No. 09/649,484	) Examiner: Sotomayor, John
Filing Date: August 28, 2000	) Group Art Unit No.: 3714
For    METHOD FOR EVALUATING	)
STANDARDIZED TEST RESULTS	)

**AFFIDAVIT UNDER 37 CFR 1.132**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Dear Sir:

In response to the Office Action mailed on March 30, 2004, Applicant submits the following affidavit:

I, Eliot R. Long, am currently self-employed as a consultant to organizations developing, distributing or using standardized test materials. I am a graduate of the Northwestern University Kellogg Graduate School of Management and the former Vice President of Research and Development at Wonderlic, Inc., a publisher/distributor of psychological testing materials. In my present capacity, I served as a consultant to the Board of Education of the City of New York and to Wonderlic, Inc. where, in both cases, I provided reviews of test administration standardization utilizing the method of the invention of the present application.

### BACKGROUND INFORMATION

The present invention relates to methods for evaluating standardized test administrations to a population of subjects to determine whether a test was properly administered to a particular sub-group, or class. Individual test results are expected to vary. The entire purpose of testing is to devise instruments (tests) on which individual performance varies and to then measure individual results against the pattern of variation established by other individuals who take the same test or against performance standards. When test developers devise their instruments, they are equally determined to eliminate variation due to non-test characteristics. Thus, for example, the reading requirements are minimized on math tests and all test takers are given the same set of test instructions. Over the length of the twentieth century, test developers have improved their art to create test content more accurately focused to its intended purpose, to construct the test materials or software to represent the content without changing its difficulty or adding additional constraints, and to specify the test instructions and administration to provide each test taker with the same, neutral opportunity to express his or her capabilities. All this with the single purpose to ensure that variation in the test results reflects individual differences on the test construct.

Just as individual, and even group, test scores are expected to vary with the capabilities of the test takers, the pattern of internal test performance that sums up to their test results is expected to be consistent. Relatively difficult questions are expected to be relatively difficult for all test takers and in all test taker groups (i.e. classrooms). The relative difficulty of each test question is reflected in the percentage of test takers who answer the question correctly. Easy question: high percent correct; difficult question: low percent correct. These question by question percentages may be assembled, over all the questions of the test, to make up a group or classroom profile. Any one profile should be unremarkable, other than the percentages, in

general, may be higher or lower depending on the skill level of the group. Nevertheless, the percentages should rise at the easier questions and fall at the more difficult questions in a similar pattern for all groups. The pattern of this profile should be consistent even when the test scores, by individual or by group, vary. This consistency underlies test score reliability. The power of this consistency, in the presence of normal, construct related variation, to set a norm and illuminate instances of improper influence is clearly unanticipated by the prior art.

In the present invention, the unit of analysis is the question-by-question pattern of group behavior under the direction of the same test administrator or subject to a common element affecting test administration. In such groups, individual behavior is subsumed to the group. Groups may vary in skill level (the construct of the test) such that higher skilled groups will achieve a higher success rate (higher percents correct) on the test and lower skilled groups a lower success rate (lower percents correct). But, the pattern of relatively higher and lower percents correct within the profile, on a question-by-question basis, will remain essentially the same for both higher and lower skill groups. Significant variation in the pattern among group profiles, on a question-by-question basis, will most likely be due to variations in behavior initiated by the test administrator, not by the students or test takers.

Thus, while the methods of organization of data or of statistical analysis may be applied to both the norms of student behavior and the norms of test administrator behavior, the behaviors are categorically different and unrelated. The great gap between them is evidenced by the large number of instances of alleged teacher cheating, the earnest efforts by experts in the field to confirm or disconfirm the allegations, and yet the failure of any prior development of the method of the Applicant's invention. The inventor has applied the method of the invention in a large number of test administration reviews. These reviews have confirmed the normal consistency of group test item response patterns (group or classroom profiles) and the sharp divergence that occurs when an improper influence has been applied.

#### FAILURE IN THE PRIOR ART

An example of a recent investigation of alleged teacher cheating

In December, 1999, the Special Commissioner of Investigation for the New York City School District, Edward F. Stancik, released the report, "Cheating the Children: Educator Misconduct on Standardized Tests" (Loughran & Comiskey, 1999). This report accused 52 New York City teachers, principals, and other school employees of improperly assisting their students to higher test scores. A torrent of highly charged publicity, some against the schools and some against the Special Commissioner, quickly followed. A year later, the teachers' union, the United Federation of Teachers (UFT), released their own report of an investigation of the Special Commissioner's methods of investigation. The UFT report strongly criticized the Commissioner's use of interviews with students and discounted the statistical evidence as "flimsy." Although all of the accused teachers and principals were initially suspended, most were reinstated, some with a letter of reprimand. A small number of teachers were found guilty and fired (these were cases where "cheat sheets" and other hard evidence was found). Yet, for the great majority, the evidence was considered insufficient.

The evidence in the Stancik report was largely that of statements from witnesses, and many of these were the students themselves. The teachers' union and the press challenged the ethics of the methods used in interviewing students and challenged the motives of many of the other witnesses. The chief 'statistical' evidence was based on analyses of the erasure patterns on student answer sheets. A previous New York City study of student erasure patterns (LeDonni, 1992) indicated that whenever more than 25% of the students in a class have 5 or more erasures which change answers from incorrect to correct, the condition may be considered exceptional, though not necessarily indicate cheating. No teacher was fired based on erasure analysis.

The primary result of the Stancik report, the teachers' union report, and the press coverage was an impression of uneasy suspicion: Suspicion that the Special Commissioner was right about at least some, if not most, of the teacher assisted cheating, but wrong about the investigative methods used; suspicion that the teacher's union investigation was more a defensive attack on the Special Commissioner than a search for truth; and suspicion that the Board of Education was not diligent in controlling the testing process.

This case of alleged teacher cheating occurred in the largest school district in the United States (1,200 schools, \$11 billion budget). The Special Commissioner's investigation was carried out with the full cooperation of the Division of Assessment and Accountability with a staff of Ph.D. level trained assessment specialists. The investigation relied on the most advanced and effective methods known to be available, and yet concluded with ambiguity and great disruption in the school district. If an improved method would have been obvious to the investigators, it surely would have been applied. In fact, following the Stancik report in March 2000, the Board of Education of the City of New York hired the inventor in a consultant capacity to apply the nascent method of the present invention to determine its efficacy. Robert Tobias, Executive Director of the Division of Assessment and Accountability commented, "I've never seen anything like it," referring to the comparison of class response pattern profiles to a normative profile. Following applications of the method of the invention to test administrations at grade levels 3 through 8, Mr. Tobias commented, "It's uncannily accurate."

Importance of effectively resolving allegations of teacher cheating with minimal disruption.

Allegations of teacher assisted cheating are unfortunately common in public schools. The Board of Education of the City of New York receives as many as 100 such allegations each year. The great majority of the allegations is unsupported and arises from misinformation, yet a significant number require the school district to investigate at some level. The number of instances of teacher cheating that rise to the level of formal reports and public attention reflect the more serious cases, usually involving a number of teachers. Examples are:

Location	Report Author
Chicago, IL	Perlman, 1985
Fairfield, CT	Lindsay, 1996
Georgia	Monsaas & Englehard, 1991; Loupe, 1998
Los Angeles, CA	Aiken, 1991
Memphis, TN	Toch & Wagner, 1992
Oahu, HA	Lawton, 1996
South Carolina	Canner, 1992

The director of testing for the Austin, Texas school district may have summed it up most succinctly:

“... teachers cheat when they administer standardized tests to students. Not all teachers, not even very many of them; but enough to make cheating a major concern to all of us who use data for decision making.” (Ligon, 1985, p1)

The history of allegations of teacher assisted cheating on classroom tests establishes the clear need for methods to confirm or disconfirm such allegations when they arise. The climate of educator and public concern for confidence in the student assessment process presents the need for a proactive method of analysis that will identify instances of improper proctor influences on student test results without having to wait for allegations to be made.

Launching an investigation into alleged teacher assisted cheating presents a daunting prospect for school districts. There will be defensive actions by the teacher(s) involved and the teachers' union, there will be concern by parents, there will be media attention and the media's independent efforts to investigate, and there will be disruption in the staff schedules and the emotional burden on the students. School administrators require strong indications of cheating before incurring these consequences of an investigation.

The methods of investigation are limited, but universally begin with written statements by all involved; interviews, evidence gathering and attempts to organize and interpret the information available. The formal literature on investigations is limited to erasure analysis (e.g., Lindsay, 1996; Qualls, 2001), evaluation of unexpected test score gains (e.g., Perlman, 1985), surveys of teachers, students and others involved in testing (e.g., Gay, 1990; Monsaas & Engelhard, 1991), and student interviews (e.g., Loughran & Comiskey, 1999). There is no record of an investigation that included an analysis of student or classroom test item response patterns other than that of erasure analysis.

Expertise of investigators.

Most investigations are conducted by the school district's assessment specialists. These staff professionals are most often formally trained in psychometrics and regularly attend and contribute to conferences on assessment issues. These conferences may be organized by their state associations or by national groups such as the American Educational Research Association and the National Council on Measurement in Education. The AERA and NCME are two of the three groups (along with the American Psychological Association) who author the *Standards for Educational and Psychological Testing* (AERA, 1999).

Investigation methodsInterviews and material evidence of cheating

As noted, investigations begin with written statements and interviews. A search is made for evidence of improper actions such as 'cheat sheets,' unusual marks in test booklets, and improper materials posted on the classroom walls or chalkboard. These efforts are rarely successful and questionable items that are found are usually subject to alternate interpretations.

Erasure analysis

The investigation will most often turn to an analysis of the erasure patterns on the students answer sheets. Some school districts, such as New York City, have established rules of thumb for exceptional erasure patterns. The school district in Fairfield, Connecticut hired forensic experts to attempt to distinguish between erasures made by young students and those likely to be made by adults (Lindsay, 1996). This case received exceptional notoriety because the school had been twice honored by the U.S. Department of Education for excellence. The case became known as "Erasergate." The investigation of erasures resulted in no proof of cheating.

This case of erasure analysis is perhaps most telling of experts in assessment looking directly at an opportunity to apply the inventor's method and failing to see it. These investigators focused on the nature of the erasures (smudgy or neat, partially or completely filled in answer bubbles, etc.) rather than on the effect of the erasures on the group's test item response.

pattern. While the smudginess of the erasures proved inconclusive, their effect on the response pattern (profile) could have been evaluated with the same conclusiveness as with DNA testing.

#### *Retesting*

Faced with a lack of effective methods to evaluate evidence of cheating directly from the tests, many school districts rely on retesting the students involved. Students who have benefited from improper teacher assistance will, it is thought, score significantly lower on retesting. Ambiguity arises, however, from the fact that the time interval between the original testing and the retesting may result in students either improving their skills or forgetting specific test content. Statistically, student measurement is always subject to "regression toward the mean," meaning the likelihood that, on retesting, students with higher scores are likely to score somewhat less and students with lower scores are likely to score somewhat higher. These retesting effects most often cause inconclusive results (i.e., Perlman, 1985).

#### *Statistical methods*

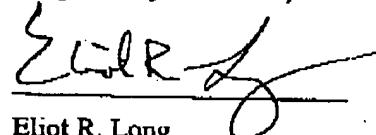
While several statistical methods (e.g., Frary, 1993; Frary & Tideman, 1997) and software packages (e.g., Advanced Psychometrics, 1993) exist for identifying student initiated cheating (all essentially based on identifying unusual matched pairs of answers indicating student copying), no such method had been introduced for identifying proctor initiated cheating as of the time of Application No. 09-649,484 (Aug. 28, 2000).

Since that date, Jacob and Levitt (2002a, 2002b) have developed a statistical method for identifying potential improper proctor influence. The Jacob and Levitt method is a further development of the analysis of unusual matched pairs of answers, looking for a high frequency of unusual pairs in classrooms with unusually large gains from the prior year's test results. The Jacob and Levitt method does not develop normative class test item profiles and measure each class against this norm. As a result, the Jacob and Levitt method is limited to only those cases where the improper proctor actions have resulted in a high frequency of exact pairs of answers, something that the inventor's research has found to represent only a limited portion of the various forms of improper practices.

Conclusion:

The instances of allegations of teacher assisted cheating identified above demonstrate a clear need for an improved method of evaluating suspect test results. These instances have occurred in environments where professionals with substantial experience and training in educational assessments are employed. The public reports of investigations into alleged teacher cheating both illustrate the disruption and pain caused to the individuals and school districts involved and the frustration and, ultimately, inconclusiveness of the investigative methods applied. Clearly, a substantial need for the benefits of the inventor's method has been present, there is a substantial history of persons with reasonable skill and experience in the art who have attempted to address the need, and yet the method of the invention has remained undiscovered until this application. Given this history, the inventor respectfully submits that the method of the invention cannot be considered to be obvious.

Respectfully submitted,



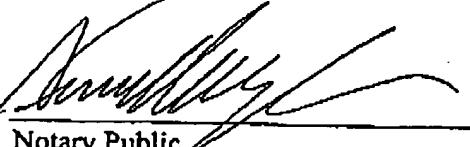
Eliot R. Long

Attached: List of References

STATE OF NY )  
COUNTY OF KING ) ss.

I, VINCENT RACCUGLIA, a Notary Public in and for the County and State aforesaid, do hereby certify that Eliot R. Long, personally known to me to be the same person whose name is subscribed to the foregoing instrument, appeared before me this day in person and acknowledged that he signed, sealed and delivered the said instrument as his free and voluntary act for the uses and purposes therein set forth.

IN WITNESS WHEREOF, I have hereunto set my hand and Notarial Seal, this 30<sup>th</sup> day of June, 2004.



Notary Public

My Commission Expires:

VINCENT R. RACCUGLIA  
Notary Public, State of New York  
No. 23-31122  
Qualified in KING County  
Commission Expires JUNE 30, 2004

*2004*

9

Commission Expires JUNE 30, 2004

## References:

Advance Psychometrics (1993). *Scrutiny!* [Computer software]. St. Paul, MN: Author

Aiken, L. R. (1991). Detecting, understanding, and controlling for cheating on tests. *Research in Higher Education*, 32(6), 725-736.

American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Canner, J. (1992). Regaining the public trust: A review of school testing programs and practices. *NASSP Bulletin*, 72(545), 6-15.

Gay, G.H. (1990). Standardized tests: Irregularities in administering of tests affect test results. *Journal of Instructional Psychology*, 17(2), 93-103.

Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education*, 6(2), 153-165.

Frary, R. B. and Tideman, T. N. (1997). Comparison of two indices of answer-copying and development of a spliced index. *Educational and Psychological Measurement*, 57(1), 20-32.

Jacob, B.A. and Levitt, S.D. (2002a). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Working Paper 9413 ([www.nber.org/papers/w9413](http://www.nber.org/papers/w9413)).

Jacob, B.A. and Levitt, S.D. (2002b). "Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory." Working Paper 9414 ([www.nber.org/papers/w9414](http://www.nber.org/papers/w9414)).

Lawton, M. (1996, Nov. 13). Alleged tampering underscores pitfalls of testing. *Education Week*, 5.

LeDonni, R. (1992). *A presentation of the results of test answer document erasure analysis*. The Board of Education of The City of New York.

Ligon, G. (1985). *Opportunity knocked out: Reducing cheating by teachers on student tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL (ERIC Document Reproduction Service No. ED 263181)

Lindsay, D. (1996, Oct. 2). Whodunit? Officials find thousands of erasures on standardized tests and suspect tampering. *Education Week*, 25-29.

Loughran, R, and Comiskey, T. (1999, December). *Cheating the Children: Educator Misconduct on Standardized Tests*. The Special Commissioner of Investigation for the New York City School District.

Loupe, D. (1998, March 24). Cheating inquiries abound on two key tests. *Atlanta Journal and Constitution*, pA1.

Monsaas, J.A., and Engelhard, G., Jr. (1991, April). *Attitudes toward testing practices as cheating and teachers' testing practices*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service, No. ED338643)

Perlman, C.L. (1985, March). *Results of a citywide testing program audit in Chicago*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service, No. ED263212)

Qualls, A. L. (2001). Can knowledge of erasure behavior be used as an indicator of possible cheating? *Educational Measurement: Issues and Practice*, 20(1), 9-16.

Toch, T. and Wagner, B. (1992, April 27) Schools for scandal. *U.S. News and World Report*, 62-72.